

## MAIZE RAR1-INTERACTOR POLYNUCLEOTIDES AND METHODS OF USE

## CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation in-part of U.S. Application No. 10/074,473, filed February 12, 2002, which claims the benefit of U.S. Provisional Application No. 60/268,157, filed February 12, 2001, each of which are incorporated in their entirety by  
5 reference herein.

## FIELD OF THE INVENTION

10 The present invention relates generally to plant molecular biology. More specifically, it relates to nucleic acids and methods for modulating their expression in plants and to transforming genes into plants in order to enhance disease resistance.

## BACKGROUND OF THE INVENTION

15 Plant disease outbreaks have resulted in catastrophic crop failures that have triggered famines and caused major social change. Generally, the best strategy for plant disease control is to use resistant cultivars selected or developed by plant breeders for this purpose. However, the potential for serious crop disease epidemics persists today, as evidenced by outbreaks of the Victoria blight of oats and southern corn leaf blight.  
20 Accordingly, molecular methods are needed to supplement traditional breeding methods to protect plants from pathogen attack.

Disease in plants is caused by biotic and abiotic causes. Biotic causes include fungi, viruses, insects, bacteria, and nematodes. Of these, fungi are the most frequent causative agents of disease in plants. Abiotic causes of disease in plants include extremes  
25 of temperature, water, oxygen, soil pH, plus nutrient-element deficiencies and imbalances, excess heavy metals, and air pollution.

A host of cellular processes enables plants to defend themselves from disease caused by pathogenic agents. These processes apparently form an integrated set of resistance mechanisms that is activated by initial infection and then limits further spread of the invading pathogenic microorganism.

5           Subsequent to recognition of a potentially pathogenic microbe, plants can activate an array of biochemical responses. Generally, the plant responds by inducing several local responses in the cells immediately surrounding the infection site. The most common resistance response observed in both nonhost and race-specific interactions is termed the "hypersensitive response" (HR). In the hypersensitive response, cells  
10       contacted by the pathogen, and often neighboring cells, rapidly collapse and dry in a necrotic fleck. Other responses include the deposition of callose, the physical thickening of cell walls by lignification, and the synthesis of various antibiotic small molecules and proteins. Genetic factors in both the host and the pathogen determine the specificity of these local responses, which can be very effective in limiting the spread of infection.

15           The hypersensitive response in many plant-pathogen interactions results from the expression of a resistance (R) gene in the plant and a corresponding avirulence (avr) gene in the pathogen. This interaction is associated with the hypersensitive response. R genes that respond to specific bacterial, fungal, or viral pathogens, have been isolated from a variety of plant species and several appear to encode cytoplasmic proteins.

20           The resistance gene in the plant and the avirulence gene in the pathogen often conform to a gene-for-gene relationship. That is, resistance to a pathogen is only observed when the pathogen carries a specific avirulence gene and the plant carries a corresponding or complementing resistance gene. Because avrR gene-for-gene relationships are observed in many plant-pathogen systems and are accompanied by a  
25       characteristic set of defense responses, a common molecular mechanism underlying avrR gene mediated resistance has been postulated. A simple model which has been proposed is that pathogen avr genes directly or indirectly generate a specific molecular signal that is recognized by cognate receptors encoded by plant R genes.

30           Both plant resistance genes and corresponding pathogen avirulence genes have been cloned. Race-specific single gene resistance controls powdery mildew, a common fungal pathogen. Such resistance is overcome when the pathogen evolves races that are

not recognized by the resistance (R) gene in use in crops. In barley, the gene Rar1 is required for the powdery mildew resistance gene Mla12 to function. Barley Rar1 has been isolated (Shirasu, *et al.*, *Cell*, 99:355-366 (1999)). Rar1 is also required for the function of some race-specific mildew resistance genes unrelated to Mla12. Shirasu, *et al.* also observed that Rar1 contains a protein domain, which they termed CHORD (Shirasu, *supra*). The CHORD domains bind  $Zn^{++}$  and are conserved in tandem organization in protozoa, plants and metazoa.

In the present invention, new polynucleotides coding for proteins, which interact with maize Rar1 have been isolated from maize. By manipulation of the Rar1-interacting polynucleotides in plants, a plant can be engineered to improve resistance to pathogens by modulating the expression of Rar1. The present invention provides a new method for altering expression of the maize Rar1 gene.

#### SUMMARY OF THE INVENTION

Generally, it is the object of the present invention to provide nucleic acids and proteins relating to Rar1-interactors. It is an object of the present invention to provide transgenic plants comprising the nucleic acids of the present invention. It is another object of the present invention to provide methods for modulating, in a transgenic plant, the expression of the nucleic acids of the present invention.

Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a) a polynucleotide amplified from a *Zea mays* nucleic acid library using the primers of the present invention; (b) a polynucleotide encoding a maize Rar1-interactor protein; (c) a polynucleotide having at least 80%, or 90% sequence identity to the polynucleotides of the present invention; (d) a polynucleotide that hybridizes under highly stringency conditions to the polynucleotides of the present invention; (e) a polynucleotide having the sequence shown in SEQ ID NOS: 1, 2, 3, 7, or 9; and (f) a polynucleotide complementary to a polynucleotide of (a) through (e). The isolated nucleic acid can be DNA. The isolated nucleic acid can also be RNA.

In another aspect, the present invention relates to vectors comprising the polynucleotides of the present invention. Also the present invention relates to

recombinant expression cassettes, comprising a nucleic acid of the present invention operably linked to a promoter.

In another aspect, the present invention is directed to a host cell into which has been introduced the recombinant expression cassette.

5 In yet another aspect, the present invention relates to a transgenic plant or plant cell comprising a recombinant expression cassette with a promoter operably linked to any of the isolated nucleic acids of the present invention. Preferred plants containing the recombinant expression cassette of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice barley, and  
10 millet. The present invention also provides transgenic seed from the transgenic plant.

In further aspect, the present invention relates to a method of modulating the level of protein in a plant by introducing into a plant cell a recombinant expression cassette comprising a polynucleotide of the present invention operably linked to a promoter; culturing the plant cell under plant growing conditions to produce a regenerated plant;  
15 and inducing expression of the polynucleotide for a time sufficient to modulate the protein of the present invention in the plant. In addition, the present invention relates to methods of enhancing disease resistance in plants. Preferred plants of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. The level of protein in the plant can either  
20 be increased or decreased.

## DETAILED DESCRIPTION OF THE INVENTION

### Overview

The present invention provides, among other things, compositions and methods  
25 for modulating (i.e., increasing or decreasing) the level of polynucleotides and polypeptides of the present invention in plants. In particular, the polynucleotides and polypeptides of the present invention can be expressed temporally or spatially, e.g., at developmental stages, in tissues, and/or in quantities, which are uncharacteristic of non-recombinantly engineered plants. Thus, the present invention provides utility in such  
30 exemplary applications as engineering disease resistance.

The present invention also provides isolated nucleic acid comprising polynucleotides of sufficient length and complementarity to a gene of the present invention to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcripts. For example, isolated nucleic acids of the present invention can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants, for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants (polymorphisms), orthologs, or paralogs of the gene, or for site directed mutagenesis in eukaryotic cells (see, e.g., U.S. patent No. 5,565,350). The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides, or for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more genes of the present invention in a host cell, tissue, or plant. Attachment of chemical agents, which bind, intercalate, cleave and/or crosslink to the isolated nucleic acids of the present invention can also be used to modulate transcription or translation. The present invention also provides isolated proteins comprising a polypeptide of the present invention (e.g., proenzyme, proenzyme, or enzymes).

The isolated nucleic acids and proteins of the present invention can be used over a broad range of plant types, particularly monocots such as the species of the family *Gramineae* including *Sorghum bicolor* and *Zea mays*. The isolated nucleic acid and proteins of the present invention can also be used in species from the genera: *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Pisum*, *Phaseolus*, *Lolium*, *Oryza*, *Avena*, *Hordeum*, *Secale*, *Allium*, and *Triticum*.

The invention is drawn to compositions and methods for inducing resistance in a plant to plant pests. Accordingly, the compositions and methods are also useful in protecting plants against fungal pathogens, viruses, nematodes, insects and the like.

By "disease resistance" is intended that the plants avoid the disease symptoms that are the outcome of plant-pathogen interactions. That is, pathogens are prevented from causing plant diseases and the associated disease symptoms, or alternatively, the disease symptoms caused by the pathogen is minimized or lessened.

By "antipathogenic compositions" is intended that the compositions of the invention have antipathogenic activity and thus are capable of suppressing, controlling, and/or killing the invading pathogenic organism. An antipathogenic composition of the invention will reduce the disease symptoms resulting from pathogen challenge by at least about 5% to about 50%, at least about 10% to about 60%, at least about 30% to about 70%, at least about 40% to about 80%, or at least about 50% to about 90% or greater. Hence, the methods of the invention can be utilized to protect plants from disease, particularly those diseases that are caused by plant pathogens.

Assays that measure antipathogenic activity are commonly known in the art, as are methods to quantitate disease resistance in plants following pathogen infection. See, for example, U.S. Patent No. 5,614,395, herein incorporated by reference. Such techniques include, measuring over time, the average lesion diameter, the pathogen biomass, and the overall percentage of decayed plant tissues. For example, a plant either expressing an antipathogenic polypeptide or having an antipathogenic composition applied to its surface shows a decrease in tissue necrosis (*i.e.*, lesion diameter) or a decrease in plant death following pathogen challenge when compared to a control plant that was not exposed to the antipathogenic composition. Alternatively, antipathogenic activity can be measured by a decrease in pathogen biomass. For example, a plant expressing an antipathogenic polypeptide or exposed to an antipathogenic composition is challenged with a pathogen of interest. Over time, tissue samples from the pathogen-inoculated tissues are obtained and RNA is extracted. The percent of a specific pathogen RNA transcript relative to the level of a plant specific transcript allows the level of pathogen biomass to be determined. See, for example, Thomma *et al.* (1998) *Plant Biology* 95:15107-15111, herein incorporated by reference.

Furthermore, *in vitro* antipathogenic assays include, for example, the addition of varying concentrations of the antipathogenic composition to paper disks and placing the disks on agar containing a suspension of the pathogen of interest. Following incubation, clear inhibition zones develop around the discs that contain an effective concentration of the antipathogenic polypeptide (Liu *et al.* (1994) *Plant Biology* 91:1888-1892, herein incorporated by reference). Additionally, microspectrophotometrical analysis can be used to measure the *in vitro* antipathogenic properties of a composition (Hu *et al.* (1997) *Plant Mol. Biol.* 34:949-959 and Cammue *et al.* (1992) *J. Biol. Chem.* 267: 2228-2233, both of which are herein incorporated by reference).

Pathogens of the invention include, but are not limited to, viruses or viroids, bacteria, insects, fungi, and the like. Viruses include tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Specific fungal and viral pathogens for the major crops include: Soybeans: *Phytophthora megasperma* fsp. *glycinea*, *Macrophomina phaseolina*, *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Fusarium oxysporum*, *Diaporthe phaseolorum* var. *sojae* (*Phomopsis sojae*), *Diaporthe phaseolorum* var. *caulivora*, *Sclerotium rolfsii*, *Cercospora kikuchii*, *Cercospora sojae*, *Peronospora manshurica*, *Colletotrichum dematium* (*Colletotrichum truncatum*), *Corynespora cassiicola*, *Septoria glycines*, *Phyllosticta sojicola*, *Alternaria alternata*, *Pseudomonas syringae* p.v. *glycinea*, *Xanthomonas campestris* p.v. *phaseoli*, *Microsphaera diffusa*, *Fusarium semitectum*, *Phialophora gregata*, Soybean mosaic virus, *Glomerella glycines*, Tobacco Ring spot virus, Tobacco Streak virus, *Phakopsora pachyrhizi*, *Pythium aphanidermatum*, *Pythium ultimum*, *Pythium debaryanum*, Tomato spotted wilt virus, *Heterodera glycines* *Fusarium solani*; Canola: *Albugo candida*, *Alternaria brassicae*, *Leptosphaeria maculans*, *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Mycosphaerella brassicicola*, *Pythium ultimum*, *Peronospora parasitica*, *Fusarium roseum*, *Alternaria alternata*; Alfalfa: *Clavibacter michiganense* subsp. *insidiosum*, *Pythium ultimum*, *Pythium irregulare*, *Pythium splendens*, *Pythium debaryanum*, *Pythium aphanidermatum*, *Phytophthora megasperma*, *Peronospora trifoliorum*, *Phoma medicaginis* var. *medicaginis*, *Cercospora medicaginis*, *Pseudopeziza medicaginis*, *Leptotrichia medicaginis*, *Fusarium-atrum*, *Xanthomonas campestris* p.v. *alfalfae*, *Aphanomyces euteiches*, *Stemphylium herbarum*, *Stemphylium alfalfae*; Wheat:

- Pseudomonas syringae* p.v. *atrofaciens*, *Urocystis agropyri*, *Xanthomonas campestris* p.v. *translucens*, *Pseudomonas syringae* p.v. *syringae*, *Alternaria alternata*, *Cladosporium herbarum*, *Fusarium graminearum*, *Fusarium avenaceum*, *Fusarium culmorum*, *Ustilago tritici*, *Ascochyta tritici*, *Cephalosporium gramineum*,
- 5 *Collotetrichum graminicola*, *Erysiphe graminis* f.sp. *tritici*, *Puccinia graminis* f.sp. *tritici*, *Puccinia recondita* f.sp. *tritici*, *Puccinia striiformis*, *Pyrenophora tritici-repentis*, *Septoria nodorum*, *Septoria tritici*, *Septoria avenae*, *Pseudocercospora herpotrichoides*, *Rhizoctonia solani*, *Rhizoctonia cerealis*, *Gaeumannomyces graminis* var. *tritici*, *Pythium aphanidermatum*, *Pythium arrhenomanes*, *Pythium ultimum*,
- 10 *Bipolaris sorokiniana*, Barley Yellow Dwarf Virus, Brome Mosaic Virus, Soil Borne Wheat Mosaic Virus, Wheat Streak Mosaic Virus, Wheat Spindle Streak Virus, American Wheat Striate Virus, *Claviceps purpurea*, *Tilletia tritici*, *Tilletia laevis*, *Ustilago tritici*, *Tilletia indica*, *Rhizoctonia solani*, *Pythium arrhenomanes*, *Pythium graminicola*, *Pythium aphanidermatum*, High Plains Virus, European wheat striate virus; Sunflower:
- 15 *Plasmophora halstedii*, *Sclerotinia sclerotiorum*, Aster Yellows, *Septoria helianthi*, *Phomopsis helianthi*, *Alternaria helianthi*, *Alternaria zinniae*, *Botrytis cinerea*, *Phoma macdonaldii*, *Macrophomina phaseolina*, *Erysiphe cichoracearum*, *Rhizopus oryzae*, *Rhizopus arrhizus*, *Rhizopus stolonifer*, *Puccinia helianthi*, *Verticillium dahliae*, *Erwinia carotovorum* p.v. *Carotovora*, *Cephalosporium acremonium*, *Phytophthora cryptogea*,
- 20 *Albugo tragopogonis*; Maize: *Fusarium moniliforme* var. *subglutinans*, *Erwinia stewartii*, *Fusarium moniliforme*, *Gibberella zeae* (*Fusarium graminearum*), *Stenocarpella maydis* (*Diplodia maydis*), *Pythium irregulare*, *Pythium debaryanum*, *Pythium graminicola*, *Pythium splendens*, *Pythium ultimum*, *Pythium aphanidermatum*, *Aspergillus flavus*, *Bipolaris maydis* O,T (*Cochliobolus heterostrophus*),
- 25 *Helminthosporium carbonum* I, II & III (*Cochliobolus carbonum*), *Exserohilum turcicum* I, II & III, *Helminthosporium pedicellatum*, *Physoderma maydis*, *Phyllosticta maydis*, *Kabat-Maydis*, *Cercospora sorghi*, *Ustilago maydis*, *Puccinia sorghi*, *Puccinia polysora*, *Macrophomina phaseolina*, *Penicillium oxalicum*, *Nigrospora oryzae*, *Cladosporium herbarum*, *Curvularia lunata*, *Curvularia inaequalis*, *Curvularia*
- 30 *pallens*, *Clavibacter michiganense* subsp. *nebraskense*, *Trichoderma viride*, Maize Dwarf Mosaic Virus A & B, Wheat Streak Mosaic Virus, Maize Chlorotic Dwarf Virus,



*Claviceps sorghi*, *Pseudomonas avenae*, *Erwinia chrysanthemi* p.v. *Zea*, *Erwinia corotovora*, *Cornstunt spiroplasma*, *Diplodia macrospora*, *Sclerophthora macrospora*, *Peronosclerospora sorghi*, *Peronosclerospora philippinesis*, *Peronosclerospora maydis*, *Peronosclerospora sacchari*, *Spacelotheca reiliana*, *Physopella zea*, *Cephalosporium*  
 5 *maydis*, *Cephalosporium acremonium*, Maize Chlorotic Mottle Virus, High Plains Virus, Maize Mosaic Virus, Maize Rayado Fino Virus, Maize Streak Virus, Maize Stripe Virus, Maize Rough Dwarf Virus; Sorghum: *Exserohilum turcicum*, *Colletotrichum graminicola* (*Glomerella graminicola*), *Cercospora sorghi*, *Gloeocercospora sorghi*, *Ascochyta sorghina*, *Pseudomonas syringae* p.v. *syringae*, *Xanthomonas campestris* p.v.  
 10 *holcicola* *Pseudomonas andropogonis*, *Puccinia purpurea*, *Macrophomina phaseolina*, *Perconia circinata*, *Fusarium moniliforme*, *Alternaria alternate*, *Bipolaris sorghicola*, *Helminthosporium sorghicola*, *Curvularia lunata*, *Phoma insidiosa*, *Pseudomonas avenae* (*Pseudomonas alboprecipitans*), *Ramulispora sorghi*, *Ramulispora sorghicola*, *Phyllachara sacchari*, *Sporisorium reilianum* (*Sphacelotheca reiliana*), *Sphacelotheca*  
 15 *cruenta*, *Sporisorium sorghi*, Sugarcane mosaic H, Maize Dwarf Mosaic Virus A & B, *Claviceps sorghi*, *Rhizoctonia solani*, *Acremonium strictum*, *Sclerophthora macrospora*, *Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Sclerospora graminicola*, *Fusarium graminearum*, *Fusarium oxysporum*, *Pythium arrhenomanes*, *Pythium graminicola*, etc.

20 Nematodes include parasitic nematodes such as root-knot, cyst, and lesion nematodes, including *Heterodera* spp., *Meloidogyne* spp., and *Globodera* spp.; particularly members of the cyst nematodes, including, but not limited to, *Heterodera glycines* (soybean cyst nematode); *Heterodera schachtii* (beet cyst nematode); *Heterodera avenae* (cereal cyst nematode); and *Globodera rostochiensis* and *Globodera*  
 25 *pailida* (potato cyst nematodes). Lesion nematodes include *Pratylenchus* spp.

Plasmids containing the nucleotide sequences of the invention, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:7, and SEQ ID NO:9 were deposited with the Patent Depository of the American Type Culture Collection (ATCC), Manassas, Virginia, on February 21, 2001 and assigned Patent Deposit No. PTA-3031. These deposits will be  
 30 maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. These deposits

were made merely as a convenience for those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112.

### Definitions

5           Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their  
10       commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

          By "amplified" is meant the construction of multiple copies of a nucleic acid  
15       sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-Beta Replicase systems, transcription-based amplification system (TAS), and strand  
20       displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, DH Persing *et al.*, Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

          As used herein, "antisense orientation" includes reference to a duplex  
25       polynucleotide sequence, which is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

          By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid  
30       encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated

sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as are present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma*  
5 *capricolum*, or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed therein.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be  
10 expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.*, *Nucl. Acids Res.* 17: 477-498 (1989)). Thus, the maize preferred codon for a particular amino acid might be derived from known gene sequences from maize. Maize  
15 codon usage for 28 genes from maize plants is listed in Table 4 of Murray *et al.*, *supra*.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene  
20 is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species, or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell, which contains a vector and supports the replication  
25 and/or expression of the vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "introduced" in the context of inserting a nucleic acid into a cell, means  
30 "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic

acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The terms "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components that normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment; or (2) if the material is in its natural environment, the material has been synthetically (non-naturally) altered by deliberate human intervention to a composition and/or placed at a location in the cell (e.g., genome or subcellular organelle) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which has been altered, by means of human intervention performed within the cell from which it originates. See, e.g., Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells, Kmiec, U.S. patent No. 5,565,350; *In Vivo* Homologous Sequence Targeting in Eukaryotic Cells; Zarling *et al.*, PCT/US93/03868. Likewise, a naturally occurring nucleic acid (e.g., a promoter) becomes isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid. Nucleic acids, which are "isolated", as defined herein, are also referred to as "heterologous" nucleic acids.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules, which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in*

*Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, *Molecular Cloning – A Laboratory Manual*, 2<sup>nd</sup> ed., Vol. 1-3 (1989); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994).

5       As used herein “operably linked” includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in  
10       the same reading frame.

          As used herein, the term “plant” includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes,  
15       pollen, and microspores. The class of plants, which can be used in the methods of the invention, is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Preferred plants include, but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. A particularly preferred plant is maize (*Zea*  
20       *mays*).

          As used herein, “polynucleotide” includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring  
25       nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons  
30       are “polynucleotides” as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to

name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modification have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including among other things, simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide", and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, this invention contemplates the use of both the methionine containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not

found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a result of deliberate human intervention. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g.,  
5 spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a host cell. The recombinant  
10 expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term "residue" or "amino acid residue" or "amino acid" are used  
15 interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

20 The term "selectively hybridizes" includes a reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least  
25 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The terms "stringent conditions" or "stringent hybridization conditions" include reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background).  
30 Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing

conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. It is known in the art that a hybridization step of 6-8 hours is preferred and a final wash step of 30 minutes up to an hour is preferred. See generally, Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989).

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the  $T_m$  can be approximated from the equation of Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284 (1984):  $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%CG) - 0.61 (\% \text{ form}) - 500/L$ ; where M is the molarity of monovalent cations, %CG is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe.  $T_m$  is reduced by about 1°C for each 1% of mismatching; thus,  $T_m$ , hybridization and/or wash conditions



can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with  $\geq 90\%$  identity are sought, the  $T_m$  can be decreased  $10^\circ\text{C}$ . Generally, stringent conditions are selected to be about  $5^\circ\text{C}$  lower than the thermal melting point ( $T_m$ ) for the specific sequence and its complement at a defined ionic strength and pH.

5 However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or  $4^\circ$  lower than the thermal melting point ( $T_m$ ); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or  $10^\circ\text{C}$  lower than the thermal melting point ( $T_m$ ); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or  $20^\circ\text{C}$  lower than the thermal melting point ( $T_m$ ). Using the equation,  
10 hybridization and wash compositions, and desired  $T_m$ , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a  $T_m$  of less than  $45^\circ\text{C}$  (aqueous solution) or  $32^\circ\text{C}$  (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the  
15 hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 “Overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995).

20 As used herein, “transgenic plant” includes reference to a plant, which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. “Transgenic” is  
25 used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term “transgenic” as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by  
30 conventional plant breeding methods or by naturally occurring events such as random

cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often  
5 replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison windows", (c) "sequence identity", and (d) "percentage of sequence identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis  
10 for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" means includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the  
15 polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can  
20 be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison may be conducted by the local alignment  
25 algorithm of Smith and Waterman. *Adv. Appl. Math.* 2: 482 (1981); by the global alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970); by the search for local alignment method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View,  
30 California, GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison,

Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73: 237-244 (1988); Higgins and Sharp, *CABIOS* 5: 151-153 (1989); Corpet, *et al.*, *Nucleic Acids Research* 16: 10881-90 (1988); Huang, *et al.*, *Computer Applications in the Biosciences* 8: 155-65 (1992), and Pearson, *et al.*, *Methods in Molecular Biology* 24: 307-331 (1994). The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995).

GAP uses the algorithm of Needleman and Wunsch (*J Mol Biol* 48: 443-453 (1970)) to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package are 8 and 2, respectively, for protein sequences. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 100. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity

threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (*see* Henikoff and Henikoff, *Proc Natl Acad Sci USA* 89:10915). Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the GAP version 10 of Wisconsin Genetic Software Package using  
5 default parameters.

Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence which either match or satisfy some  
10 positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased.

15 Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always <0) . For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum  
20 achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4,  
25 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89: 10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin &  
30 Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides

an indication of the probability that a match between two nucleotide or two amino acid sequences would occur by chance.

BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences, which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, *Comput. Chem.*, 17:149-163 (1993)) and XNU (Claverie and States, *Comput. Chem.*, 17:191-201 (1993)) low-complexity filters can be employed alone or in combination.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences, which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences, which differ by such conservative substitutions, are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, *Computer Applic. Biol. Sci.*, 4: 11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

### **Nucleic Acids**

The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

A polynucleotide of the present invention is inclusive of:

- (a) a polynucleotide encoding a polypeptide of the present invention, including exemplary polynucleotides of SEQ ID NOS: 1, 2, 3, 7, or 9;
- (b) a polynucleotide which is the product of amplification from a *Zea mays* nucleic acid library using primer pairs which selectively hybridize under stringent conditions to loci within a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 2, 3, 7, or 9;
- (c) a polynucleotide which selectively hybridizes to a polynucleotide of (a) or (b);
- (d) a polynucleotide having a specified sequence identity with polynucleotides of (a), (b), or (c);
- (e) complementary sequences of polynucleotides of (a), (b), (c), or (d);
- (f) a polynucleotide comprising at least a specific number of contiguous nucleotides from a polynucleotide of (a), (b), (c), (d), or (e); and
- (g) an isolated polynucleotide made by the process of: 1) providing a full-length enriched nucleic acid library, 2) selectively hybridizing the polynucleotide to a

polynucleotide of (a), (b), (c), (d), (e), (f), (g), or (h), thereby isolating the polynucleotide from the nucleic acid library. The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

5

*A. Polynucleotides Encoding A Polypeptide of the Present Invention*

The present invention provides isolated nucleic acids comprising a polynucleotide of the present invention, wherein the polynucleotide encodes a polypeptide of the present invention. Every nucleic acid sequence herein that encodes a polypeptide also, by  
10 reference to the genetic code, describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Thus, each silent variation of a nucleic acid, which encodes a polypeptide of the present  
15 invention, is implicit in each described polypeptide sequence and is within the scope of the present invention. Accordingly, the present invention includes polynucleotides of the present invention and polynucleotides encoding a polypeptide of the present invention.

*B. Polynucleotides Amplified from a Plant Nucleic Acid Library*

20 The present invention provides an isolated nucleic acid comprising a polynucleotide of the present invention, wherein the polynucleotides are amplified, under nucleic acid amplification conditions, from a plant nucleic acid library. Nucleic acid amplification conditions for each of the variety of amplification methods are well known to those of ordinary skill in the art. The plant nucleic acid library can be constructed  
25 from a monocot such as a cereal crop. Exemplary cereals include corn, sorghum, alfalfa, canola, wheat, or rice. The plant nucleic acid library can also be constructed from a dicot such as soybean. *Zea mays* lines B73, PHRE1, A632, BMS-P2#10, W23, and Mo17 are known and publicly available. Other publicly known and available maize lines can be  
30 obtained from the Maize Genetics Cooperation (Urbana, IL). Wheat lines are available from the Wheat Genetics Resource Center (Manhattan, KS).

The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. cDNA libraries can be normalized to increase the representation of relatively rare cDNAs. In optional embodiments, the cDNA library is constructed using an enriched full-length  
5 cDNA synthesis method. Examples of such methods include Oligo-Capping (Maruyama, K. and Sugano, S. *Gene* 138: 171-174, 1994), Biotinylated CAP Trapper (Carninci, *et al. Genomics* 37: 327-336, 1996), and CAP Retention Procedure (Edery, E., Chu, L.L., *et al. Molecular and Cellular Biology* 15: 3363-3371, 1995). Rapidly growing tissues or rapidly dividing cells are preferred for use as a mRNA source for construction of a cDNA  
10 library. Growth stages of corn is described in "How a Corn Plant Develops," Special Report No. 48, Iowa State University of Science and Technology Cooperative Extension Service, Ames, Iowa, Reprinted February 1993.

A polynucleotide of this embodiment (or subsequences thereof) can be obtained, for example, by using amplification primers which are selectively hybridized and primer  
15 extended, under nucleic acid amplification conditions, to at least two sites within a polynucleotide of the present invention, or to two sites within the nucleic acid which flank and comprise a polynucleotide of the present invention, or to a site within a polynucleotide of the present invention and a site within the nucleic acid which comprises it. Methods for obtaining 5' and/or 3' ends of a vector insert are well known  
20 in the art. See, e.g., RACE (Rapid Amplification of Complementary Ends) as described in Frohman, M. A., in PCR Protocols: A Guide to Methods and Applications, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, Inc., San Diego), pp. 28-38 (1990)); see also, U.S. Pat. No. 5,470,722, and *Current Protocols in Molecular Biology*, Unit 15.6, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New  
25 York (1995); Frohman and Martin, *Techniques* 1:165 (1989).

Optionally, the primers are complementary to a subsequence of the target nucleic acid which they amplify but may have a sequence identity ranging from about 85% to 99% relative to the polynucleotide sequence which they are designed to anneal to. As those skilled in the art will appreciate, the sites to which the primer pairs will selectively  
30 hybridize are chosen such that a single contiguous nucleic acid can be formed under the desired nucleic acid amplification conditions. The primer length in nucleotides is



selected from the group of integers consisting of from at least 15 to 50. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, or 50 nucleotides in length. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence. A non-annealing sequence at the 5' end of a primer (a "tail") can be added, for example, to introduce a cloning site at the terminal ends of the amplicon.

The amplification products can be translated using expression systems well known to those of skill in the art. The resulting translation products can be confirmed as polypeptides of the present invention by, for example, assaying for the appropriate catalytic activity (e.g., specific activity and/or substrate specificity), or verifying the presence of one or more linear epitopes, which are specific to a polypeptide of the present invention. Methods for protein synthesis from PCR derived templates are known in the art and available commercially. See, e.g., Amersham Life Sciences, Inc, Catalog '97, p.354.

15

C. *Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)*

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides selectively hybridize, under selective hybridization conditions, to a polynucleotide of section (A) or (B) as discussed above. Thus, the polynucleotides of this embodiment can be used for isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). For example, polynucleotides of the present invention can be used to identify, isolate, or amplify partial or full-length clones in a deposited library. In some embodiments, the polynucleotides are genomic or cDNA sequences isolated or otherwise complementary to a cDNA from a dicot or monocot nucleic acid library. Exemplary species of monocots and dicots include, but are not limited to: maize, canola, soybean, cotton, wheat, sorghum, sunflower, alfalfa, oats, sugar cane, millet, barley, and rice. The cDNA library comprises at least 50% to 95% full-length sequences (for example, at least 50%, 60%, 70%, 80%, 90%, or 95% full-length sequences). The cDNA libraries can be normalized to increase the representation of rare sequences. See, e.g., U.S. Patent No. 5,482,845. Low stringency hybridization conditions are typically, but not exclusively, employed

with sequences having a reduced sequence identity relative to complementary sequences. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% to 80% sequence identity and can be employed to identify orthologous or paralogous sequences.

*D. Polynucleotides Having a Specific Sequence Identity with the Polynucleotides of (A), (B) or (C)*

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides have a specified identity at the nucleotide level to a polynucleotide as disclosed above in sections (A), (B), or (C), above. Identity can be calculated using, for example, the BLAST or GAP algorithms under default conditions. The percentage of identity to a reference sequence is at least 60% and, rounded upwards to the nearest integer, can be expressed as an integer selected from the group of integers consisting of from 60 to 99. Thus, for example, the percentage of identity to a reference sequence can be at least 70%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%.

*E. Polynucleotides Encoding a Protein Having a Subsequence from a Prototype Polypeptide and Cross-Reactive to the Prototype Polypeptide*

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides encode a protein having a subsequence of contiguous amino acids from a prototype polypeptide of the present invention such as are provided in section (A), above. The subsequences of a nucleotide sequence may encode protein fragments that retain the biological activity of the native protein and hence modulate disease resistance. Furthermore, although the present invention is not bound by any mechanism of action, such fragments retaining the biological activity of the native protein may interact with Rar1. Alternatively, subsequences of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, subsequences of a

nucleotide sequence may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the full-length nucleotide sequence encoding the proteins of the invention.

5 The length of contiguous amino acids from the prototype polypeptide is selected from the group of integers consisting of from at least 10 to the number of amino acids within the prototype sequence. Thus, for example, the polynucleotide can encode a polypeptide having a Rar1-interactor subsequence having at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or more contiguous amino acids from the prototype polypeptide. Further, the number of such subsequences encoded by a polynucleotide of the instant embodiment can  
10 be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

Thus, a subsequence of a Rar1-interactor nucleotide sequence may encode a biologically active portion of a Rar1-interactor protein, or it may be a fragment that can  
15 be used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of a Rar1-interactor protein can be prepared by isolating a portion of one of the a Rar1-interactor nucleotide sequences of the invention, expressing the encoded portion of the a Rar1-interactor protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the Rar1-interactor protein.  
20 Nucleic acid molecules that are subsequences of a Rar1-interactor nucleotide sequence comprise at least 16, 20, 50, 75, 100, 150, 200, 250, 300, or 350 nucleotides, or up to the number of nucleotides present in a full-length a Rar1-interactor nucleotide sequence disclosed herein (for example, 691 nucleotides for SEQ ID NO:1, 571 nucleotides for SEQ ID NO:2, 388 nucleotides for SEQ ID NO:3, 1455 nucleotides of SEQ ID NO:7, or  
25 1086 nucleotides for SEQ ID NO:9, respectively).

The proteins encoded by polynucleotides of this embodiment, when presented as an immunogen, elicit the production of polyclonal antibodies which specifically bind to a prototype polypeptide such as (but not limited to) a polypeptide encoded by the polynucleotide of sections (A) or (B) above. Generally, however, a protein encoded by a  
30 polynucleotide of this embodiment does not bind to antisera raised against the prototype polypeptide when the antisera has been fully immunosorbed with the prototype

polypeptide. Methods of making and assaying for antibody binding specificity/affinity are well known in the art. Exemplary immunoassay formats include ELISA, competitive immunoassays, radioimmunoassays, Western blots, indirect immunofluorescent assays and the like.

5           In a preferred assay method, fully immunosorbed and pooled antisera that is elicited to the prototype polypeptide can be used in a competitive binding assay to test the protein. The concentration of the prototype polypeptide required to inhibit 50% of the binding of the antisera to the prototype polypeptide is determined. If the amount of the protein required to inhibit binding is less than twice the amount of the prototype protein,  
10       then the protein is said to specifically bind to the antisera elicited to the immunogen. Accordingly, the proteins of the present invention embrace allelic variants, conservatively modified variants, and minor recombinant modifications to a prototype polypeptide.

          A polynucleotide of the present invention optionally encodes a protein having a molecular weight of the non-glycosylated protein within 20% of the molecular weight of  
15       the full-length non-glycosylated polypeptides of the present invention. Molecular weight can be readily determined by SDS-PAGE under reducing conditions. Optionally, the molecular weight is within 15% of a full-length polypeptide of the present invention, more preferably within 10% or 5%, and most preferably within 3%, 2%, or 1% of a full-length polypeptide of the present invention.

20           Optionally, the polynucleotides of this embodiment will encode a protein having a specific enzymatic activity at least 50%, 60%, 70%, 80%, or 90% of a cellular extract comprising the native, endogenous full-length polypeptide of the present invention. Further, the proteins encoded by polynucleotides of this embodiment will optionally have a substantially similar affinity constant ( $K_m$ ) and/or catalytic activity (*i.e.*, the  
25       microscopic rate constant,  $k_{cat}$ ) as the native endogenous, full-length protein. Those of skill in the art will recognize that  $k_{cat}/K_m$  value determines the specificity for competing substrates and is often referred to as the specificity constant. Proteins of this embodiment can have a  $k_{cat}/K_m$  value at least 10% of a full-length polypeptide of the present invention as determined using the endogenous substrate of that polypeptide. Optionally, the  $k_{cat}/K_m$   
30       value will be at least 20%, 30%, 40%, 50%, and most preferably at least 60%, 70%, 80%, 90%, or 95% the  $k_{cat}/K_m$  value of the full-length polypeptide of the present invention.

Determination of  $k_{cat}$ ,  $K_m$ , and  $k_{cat}/K_m$  can be determined by any number of means well known to those of skill in the art. For example, the initial rates (*i.e.*, the first 5% or less of the reaction) can be determined using rapid mixing and sampling techniques (*e.g.*, continuous-flow, stopped-flow, or rapid quenching techniques), flash photolysis, or relaxation methods (*e.g.*, temperature jumps) in conjunction with such exemplary methods of measuring as spectrophotometry, spectrofluorimetry, nuclear magnetic resonance, or radioactive procedures. Kinetic values are conveniently obtained using a Lineweaver-Burk or Eadie-Hofstee plot.

10 *F. Polynucleotides Complementary to the Polynucleotides of (A)-(E)*

The present invention provides isolated nucleic acids comprising polynucleotides complementary to the polynucleotides of paragraphs A-E, above. As those of skill in the art will recognize, complementary sequences base pair throughout the entirety of their length with the polynucleotides of sections (A)-(E) (*i.e.*, have 100% sequence identity over their entire length). Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

*G. Polynucleotides That are Subsequences of the Polynucleotides of (A)-(F)*

20 The present invention provides isolated nucleic acids comprising polynucleotides which comprise at least 15 contiguous bases from the polynucleotides of sections (A) (B), (C), (D), (E), or (F) (*i.e.*, sections (A)-(F), as discussed above). A subsequence of a Rar1-interactor nucleotide sequence may encode a biologically active portion of a Rar1-interactor protein, or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed elsewhere herein. Subsequences of a Rar1-interactor nucleotide sequence that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of a Rar1-interactor protein.

The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1400, 1450, or 1455 contiguous nucleotides in length from the polynucleotides of sections (A) through (F). Optionally, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 1000, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, or 1400 nucleotides.

Subsequences can be made by *in vitro* synthetic, *in vitro* biosynthetic, or *in vivo* recombinant methods. In optional embodiments, subsequences can be made by nucleic acid amplification. For example, nucleic acid primers will be constructed to selectively hybridize to a sequence (or its complement) within, or co-extensive with, the coding region.

The subsequences of the present invention can comprise structural characteristics of the sequence from which it is derived. Alternatively, the subsequences can lack certain structural characteristics of the larger sequence from which it is derived such as a poly (A) tail. Optionally, a subsequence from a polynucleotide encoding a polypeptide having at least one linear epitope in common with a prototype polypeptide sequence as provided in (a), above, may encode an epitope in common with the prototype sequence. Alternatively, the subsequence may not encode an epitope in common with the prototype sequence but can be used to isolate the larger sequence by, for example, nucleic acid hybridization with the sequence from which it is derived. Subsequences can be used to modulate or detect gene expression by introducing into the subsequences compounds which bind, intercalate, cleave and/or crosslink to nucleic acids. Exemplary compounds include acridine, psoralen, phenanthroline, naphthoquinone, daunomycin or chloroethylaminoaryl conjugates.

H. *Polynucleotides that are Variants of the Polynucleotides of (A)-(G).*

By "variants" is intended substantially similar sequences. For nucleotide sequences, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the Rar1-interactor polypeptides of the invention. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis, but which still encode a Rar1-interactor protein of the invention. Generally, variants of a particular nucleotide sequence of the invention will have at least about 40%, 50%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to that particular nucleotide sequence as determined by sequence alignment programs described elsewhere herein using default parameters.

Variants of a particular nucleotide sequence of the invention (i.e., the reference nucleotide sequence) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant nucleotide sequence and the polypeptide encoded by the reference nucleotide sequence. Thus, for example, isolated nucleic acids that encode a polypeptide with a given percent sequence identity to the polypeptide of SEQ ID NO:8 are disclosed. Percent sequence identity between any two polypeptides can be calculated using sequence alignment programs described elsewhere herein using default parameters. Where any given pair of polynucleotides of the invention is evaluated by comparison of the percent sequence identity shared by the two polypeptides they encode, the percent sequence identity between the two encoded polypeptides is at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity.

30

*I. Polynucleotides From a Full-length Enriched cDNA Library Having the Physico-Chemical Property of Selectively Hybridizing to a Polynucleotide of (A)-(G)*

The present invention provides an isolated polynucleotide from a full-length enriched cDNA library having the physico-chemical property of selectively hybridizing to a polynucleotide of paragraphs (A), (B), (C), (D), (E), (F), or (G) as discussed above. Methods of constructing full-length enriched cDNA libraries are known in the art and discussed briefly below. The cDNA library comprises at least 50% to 95% full-length sequences (for example, at least 50%, 60%, 70%, 80%, 90%, or 95% full-length sequences). The cDNA library can be constructed from a variety of tissues from a monocot or dicot at a variety of developmental stages. Exemplary species include maize, wheat, rice, canola, soybean, cotton, sorghum, sunflower, alfalfa, oats, sugar cane, millet, barley, and rice. Methods of selectively hybridizing, under selective hybridization conditions, a polynucleotide from a full-length enriched library to a polynucleotide of the present invention are known to those of ordinary skill in the art. Any number of stringency conditions can be employed to allow for selective hybridization. In optional embodiments, the stringency allows for selective hybridization of sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 98% sequence identity over the length of the hybridized region. Full-length enriched cDNA libraries can be normalized to increase the representation of rare sequences.

20

*J. Polynucleotide Products Made by an cDNA Isolation Process*

The present invention provides an isolated polynucleotide made by the process of: 1) providing a full-length enriched nucleic acid library, 2) selectively hybridizing the polynucleotide to a polynucleotide of paragraphs (A), (B), (C), (D), (E), (F), (G, or (H) as discussed above, and thereby isolating the polynucleotide from the nucleic acid library. Full-length enriched nucleic acid libraries are constructed as discussed in paragraph (G) and below. Selective hybridization conditions are as discussed in paragraph (G). Nucleic acid purification procedures are well known in the art. Purification can be conveniently accomplished using solid-phase methods; such methods are well known to those of skill in the art and kits are available from commercial suppliers such as Advanced Biotechnologies (Surrey, UK). For example, a polynucleotide of paragraphs (A)-(H) can

30



be immobilized to a solid support such as a membrane, bead, or particle. See, e.g., U.S. Patent No. 5,667,976. The polynucleotide product of the present process is selectively hybridized to an immobilized polynucleotide and the solid support is subsequently isolated from non-hybridized polynucleotides by methods including, but not limited to, centrifugation, magnetic separation, filtration, electrophoresis, and the like.

### **Construction of Nucleic Acids**

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot.

The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexahistidine marker sequence provides a convenient means to purify the proteins of the present invention. A polynucleotide of the present invention can be attached to a vector, adapter, or linker for cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention less the length of its polynucleotide of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors, expression vectors, adapters, and linkers is well known and extensively described in the art. For a description of various nucleic acids see, for example, Stratagene Cloning Systems, Catalogs 1999 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog '99 (Arlington Heights, IL).

30

A. *Recombinant Methods for Constructing Nucleic Acids*

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some  
5   embodiments, oligonucleotide probes, which selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired sequence in a cDNA or genomic DNA library. Isolation of RNA and construction of cDNA and genomic libraries is well known to those of ordinary skill in the art. See, e.g.,  
10   *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995).

A1. *Full-length Enriched cDNA Libraries*

A number of cDNA synthesis protocols have been described which provide  
15   enriched full-length cDNA libraries. Enriched full-length cDNA libraries are constructed to comprise at least 600%, and more preferably at least 70%, 80%, 90% or 95% full-length inserts amongst clones containing inserts. The length of insert in such libraries can be at least 2,3, 4, 5, 6, 7, 8, 9, 10 or more kilobase pairs. Vectors to accommodate inserts of these sizes are known in the art and available commercially. See, e.g.,  
20   Stratagene's lambda ZAP Express (cDNA cloning vector with 0 to 12 kb cloning capacity). An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci et al., *Genomics*, 37:327-336 (1996). Other methods for producing full-length libraries are known in the art. See, e.g., Edery et al., *Mol. Cell Biol.*, 15(6):3363-3371 (1995); and, PCT Application WO 96/34981.

25

A2 *Normalized or Subtracted cDNA Libraries*

A non-normalized cDNA library represents the mRNA population of the tissue it was made from. Since unique clones are out-numbered by clones derived from highly expressed genes their isolation can be laborious. Normalization of a cDNA library is the  
30   process of creating a library in which each clone is more equally represented. Construction of normalized libraries is described in Ko, *Nucl. Acids. Res.*, 18(19):5705-

5711 (1990); Patanjali *et al.*, *Proc. Natl. Acad. U.S.A.*, 88:1943-1947 (1991); U.S. Patents 5,482,685, 5,482,845, and 5,637,685. In an exemplary method described by Soares *et al.*, normalization resulted in reduction of the abundance of clones from a range of four orders of magnitude to a narrow range of only 1 order of magnitude. *Proc. Natl. Acad. Sci. USA*, 91:9228-9232 (1994).

Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. In this procedure, cDNA prepared from one pool of mRNA is depleted of sequences present in a second pool of mRNA by hybridization. The cDNA:mRNA hybrids are removed and the remaining un-hybridized cDNA pool is enriched for sequences unique to that pool. See, Foote *et al.* in, *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); Kho and Zarbl, *Technique*, 3(2):58-63 (1991); Sive and St. John, *Nucl. Acids Res.*, 16(22):10937 (1988); *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); and, Swaroop *et al.*, *Nucl. Acids Res.*, 19(17):4725-4730 (1991). Several cDNA subtraction kits are commercially available. See, e.g., PCR-Select (Clontech, Palo Alto, CA).

To construct genomic libraries, large segments of genomic DNA are generated by fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. Methodologies to accomplish these ends and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate molecular biological techniques and instructions sufficient to direct persons of skill through many construction, cloning, and screening methodologies are found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989), Methods in Enzymology, Vol. 152: *Guide to Molecular Cloning Techniques*, Berger and Kimmel, Eds., San Diego: Academic Press, Inc. (1987), *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Kits for construction of genomic libraries are also commercially available.

The cDNA or genomic library can be screened using a probe based upon the sequence of a polynucleotide of the present invention such as those disclosed herein.

Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent.

5           The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic  
10 acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

          PCR-based screening methods have been described. Wilfinger *et al.* describe a  
15 PCR-based method in which the longest cDNA is identified in the first step so that incomplete clones can be eliminated from study. *BioTechniques*, 22(3): 481-486 (1997). Such methods are particularly effective in combination with a full-length cDNA construction methodology, above.

#### 20    B.     *Synthetic Methods for Constructing Nucleic Acids*

          The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68: 90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68: 109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*,  
25 *Tetra. Lett.* 22: 1859-1862 (1981); the solid phase phosphoramidite triester method described by Beaucage and Caruthers, *Tetra. Letts.* 22(20): 1859-1862 (1981), *e.g.*, using an automated synthesizer, *e.g.*, as described in Needham-VanDevanter *et al.*, *Nucleic Acids Res.*, 12: 6159-6168 (1984); and, the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide.  
30 This may be converted into double stranded DNA by hybridization with a complementary sequence or by polymerization with a DNA polymerase using the single strand as a

template. One of skill will recognize that while chemical synthesis of DNA is best employed for sequences of about 100 bases or less, longer sequences may be obtained by the ligation of shorter sequences.

## 5    **Proteins**

The isolated proteins of the present invention comprise a polypeptide having at least 10 amino acids encoded by any one of the polynucleotides of the present invention as discussed more fully, above, or polypeptides which are conservatively modified variants thereof. The proteins of the present invention or variants thereof can comprise  
10    any number of contiguous amino acid residues from a polypeptide of the present invention, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length polypeptide of the present invention. Optionally, this subsequence of contiguous amino acids is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, or 40 amino acids in  
15    length, often at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids in length.

By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids  
20    at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native protein, that is, a Rar1-interactor activity as described herein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a native Rar1-interactor protein of the  
25    invention will have at least about 40%, 50%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs described elsewhere herein using default parameters. A biologically active variant of a protein of  
30    the invention may differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

As those of skill will appreciate, the present invention includes catalytically active polypeptides of the present invention (*i.e.*, enzymes). Catalytically active polypeptides have a specific activity of at least 20%, 30%, or 40%, and preferably at least 50%, 60%, or 70%, and most preferably at least 80%, 90%, or 95% that of the native (non-synthetic), endogenous polypeptide. Further, the substrate specificity ( $k_{cat}/K_m$ ) is optionally substantially similar to the native (non-synthetic), endogenous polypeptide. Typically, the  $K_m$  will be at least 30%, 40%, or 50%, that of the native (non-synthetic), endogenous polypeptide; and more preferably at least 60%, 70%, 80%, or 90%. Methods of assaying and quantifying measures of enzymatic activity and substrate specificity ( $k_{cat}/K_m$ ), are well known to those of skill in the art.

Generally, the proteins of the present invention will, when presented as an immunogen, elicit production of an antibody specifically reactive to a polypeptide of the present invention. Further, the proteins of the present invention will not bind to antisera raised against a polypeptide of the present invention which has been fully immunosorbed with the same polypeptide. Immunoassays for determining binding are well known to those of skill in the art. A preferred immunoassay is a competitive immunoassay as discussed *infra*. Thus, the proteins of the present invention can be employed as immunogens for constructing antibodies immunoreactive to a protein of the present invention for such exemplary utilities as immunoassays or protein purification techniques.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the Rar1-interactor proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.),

herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferable.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired Rar1-interactor activity. Obviously, the mutations that will be made in the DNA encoding the variant must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the protein sequence[s] encompassed herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. That is, the activity can be evaluated by non-denaturing PAGE or co-immunoprecipitation, etc. See, for example, Sambrook *et al.* and Ausubel *et al.*, (above), and Harlow *et al.*, *Antibodies – A Laboratory Manual*, (1989), herein incorporated by reference.

As discussed elsewhere herein, variant nucleotide sequences and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different Rar1-interactor coding sequences can be manipulated to create a new Rar1-interactor protein possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*.

### **Expression of Proteins in Host Cells**

Using the nucleic acids of the present invention, one may express a protein of the present invention in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian, or preferably plant cells. The cells produce the protein in a non-natural

condition. (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the present invention. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of isolated nucleic acids encoding a protein of the present invention will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter (which is either constitutive or regulatable), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding a protein of the present invention. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. One of skill would recognize that modifications could be made to a protein of the present invention without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (e.g., poly His) placed on either terminus to create conveniently located purification sequences. Restriction sites or termination codons can also be introduced.

25

#### A. *Expression in Prokaryotes*

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding sequences, include such commonly used promoters as the beta

30



lactamase (penicillinase) and lactose (lac) promoter systems (Chang et al., Nature 198:1056 (1977)), the tryptophan (trp) promoter system (Goeddel et al., Nucleic Acids Res. 8:4057 (1980)) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake et al., Nature 292:128(1981)). The inclusion of selection markers in DNA vectors transfected in *E coli*. is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

The vector is selected to allow introduction into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva et al., *Gene* 22: 229-235 (1983); Mosbach, et al., *Nature* 302:543-545 (1983)).

#### 15 B. Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, a polynucleotide of the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

Synthesis of heterologous proteins in yeast is well known. Sherman, F., et al., *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory (1982) is a well recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeasts for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

30 A protein of the present invention, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysate. The

monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay of other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect, or plant origin. Illustrative cell cultures useful for the production of the peptides are mammalian cells. Mammalian cell systems often will be in the form of minelayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (e.g. the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen *et al.*, *Immunol. Rev.* 89:49 (1986)), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (e.g., an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present invention are available, for instance, from the American Type Culture Collection.

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (See, Schneider, *J. Embryol. Exp. Morphol.* 27:353-365 (1987).

As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague, *et al.*, *J. Virol.* 45:773-781 (1983)). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector in *DNA Cloning Vol. II a Practical Approach*, D.M. Glover, Ed., IRL Press, Arlington, Virginia pp. 213-238(1985).

## Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied.

- 5 Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method, which provides for effective transformation/transfection may be employed.

### 10 A. Plant Transformation

- The genes of the present invention can be used to transform any plant. In this manner, genetically modified plants, plant cells, plant tissue, seed, and the like can be obtained. Transformation protocols may vary depending on the type of plant cell, i.e. monocot or dicot, targeted for transformation. Suitable methods of transforming plant
- 15 cells include microinjection (Crossway *et al.*, (1986) *BioTechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium* mediated transformation (Hinchey *et al.*, (1988) *Biotechnology* 6:915-921; U.S. patent No. 5,981,840 (maize); U.S. patent No. 5,932,782 (sunflower), European patent No. 0486233 (sunflower); PCT application number WO 98/49332 (sorghum)),
- 20 direct gene transfer (Paszkowski *et al.*, (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford *et al.*, U.S. Patent 4,945,050; Tomes *et al.*, "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" In Gamborg and Phillips (Eds.) *Plant Cell, Tissue and Organ Culture: Fundamental Methods*, Springer-Verlag, Berlin (1995); McCabe *et al.*, (1988) *Biotechnology* 6:923-
- 25 926); U.S. patent number 5,990,387 (maize), U.S. patent number 5,886,244 (maize); U.S. patent number 5,322,783 (sorghum)). Also see, Weissinger *et al.*, (1988) *Annual Rev. Genet.* 22:421-477; Sanford *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.*, (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe *et al.*, (1988) *Bio/Technology* 6:923-926 (soybean); Datta *et al.*, (1990) *Biotechnology* 8:736-
- 30 740 (rice); Klein *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.*, (1988) *Biotechnology* 6:559-563 (maize); Tomes *et al.*, "Direct DNA Transfer into

Intact Plant Cells via Microprojectile Bombardment" in Gamborg and Phillips (Eds.) *Plant Cell, Tissue and Organ Culture: Fundamental Methods*, Springer-Verlag, Berlin (1995) (maize); Klein *et al.*, (1988) *Plant Physiol.* 91:440-444 (maize) Fromm *et al.*, (1990) *Biotechnology* 8:833-839 (maize); Hooydaas-Van Slogteren & Hooykaas (1984) *Nature (London)* 311:763-764; Bytebier *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet *et al.*, (1985) In *The Experimental Manipulation of Ovule Tissues* ed. G.P. Chapman *et al.*, pp. 197-209. Longman, NY (pollen); Kaeppler *et al.*, (1990) *Plant Cell Reports* 9:415-418; and Kaeppler *et al.*, (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.*, (1992) *Plant Cell* 4:1495-1505 (electroporation); LI *et al.*, (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

The cells, which have been transformed, may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports*, 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristics is stably maintained and inherited and then seeds harvested to ensure the desired phenotype or other property has been achieved. One of skill will recognize that after the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of number of standard breeding techniques can be used, depending upon the species to be crossed.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plans that would produce the selected phenotype.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered expression of a polynucleotide of the present invention relative to a control plant (i.e., native, non-transgenic). Backcrossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

#### *B. Transfection of Prokaryotes, Lower Eukaryotes, and Animal Cells*

Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transfection by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextrin, electroporation, biolistics, and micro-injection of the DNA directly into the cells. The transfected cells are cultured by means well known in the art. Kuchler, R.J., *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson and Ross, Inc (1997).

#### **Modulating Polypeptide Levels and/or Composition**

The present invention further provides a method for modulating (i.e., increasing or decreasing) the concentration or composition of the polypeptides of the present invention in a plant or part thereof. Increasing or decreasing the concentration and/or the composition (i.e., the ratio of the polypeptides of the present invention) in a plant can effect modulation. The method comprised introducing into a plant cell with a

recombinant expression cassette comprising a polynucleotide of the present invention as described above to obtain a transformed plant cell, culturing the transformed plant cell under plant cell growing conditions, and inducing or repressing expression of a polynucleotide of the present invention in the plant for a time sufficient to modulate concentration and/or composition in the plant or plant part.

In some embodiments, the content and/or composition of polypeptides of the present invention in a plant may be modulated by altering, *in vivo* or *in vitro*, the promoter of a gene to up- or down- regulate gene expression. In some embodiments, the coding regions of native genes of the present invention can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868. And in some embodiments, an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the promoter operably linked to a polynucleotide of the present invention is selected for by means known to those of skill in the art such as, but not limited to, Southern blot, DNA sequencing, or PCR analysis using primers specific to the promoter and to the gene and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate the concentration and/or composition of polypeptides of the present invention in the plant. Plant forming conditions are well known in the art and discussed briefly, *supra*.

In general, concentration or composition is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned recombinant expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development. Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail, *supra*. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds, which activate expression from these promoters, are

well known in the art. In preferred embodiments, the polypeptides of the present invention are modulated in monocots, particularly maize.

### **Molecular Markers**

5           The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Optionally, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic  
10 relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Chapter 7, Clark, Ed., Springer-Verlag, Berlin (1997). For molecular marker methods, see generally, *The DNA Revolution* by Andrew H. Paterson 1996 (Chapter 2) in:  
15 *Genome Mapping in plants* (ed. Andrew H. Paterson) by Academic Press/R.G. Lands Company, Austin, Texas, pp. 7-21.

          The particular method of genotyping in the present invention may employ any number of molecular marker analytic techniques such as, but not limited to, restriction fragment length polymorphism's (RFLPs). RFLPs are the product of allelic differences  
20 between DNA restriction fragments resulting from nucleotide sequence variability. As is well known to those of skill in the art, RFLPs are typically detected by extraction of genomic DNA and digestion with a restriction enzyme. Generally, the resulting fragments are separated according to size and hybridized with a probe; single copy probes are preferred. Restriction fragments from homologous chromosomes are revealed.  
25 Differences in fragment size among alleles represent an RFLP. Thus, the present invention further provides a means to follow segregation of a gene or nucleic acid of the present invention as well as chromosomal sequences genetically linked to these genes or nucleic acids using such techniques as RFLP analysis. Linked chromosomal sequences are within 50 centiMorgans (cM), often within 40 or 30 cM, preferably within 20 or 10  
30 cM, more preferably within 5, 3, 2, or 1 cM of a gene of the present invention.

In the present invention, the nucleic acid probes employed for molecular marker mapping of plant nuclear genomes selectively hybridize, under selective hybridization conditions, to a gene encoding a polynucleotide of the present invention. In preferred embodiments, the probes are selected from polynucleotides of the present invention.

5 Typically, these probes are cDNA probes or restriction enzyme treated (e.g., *PST I*) genomic clones. The length of the probes is discussed in greater detail, *supra*, but is typically at least 15 bases in length, more preferably at least 20, 25, 30, 35, 40, or 50 bases in length. Generally, however, the probes are less than about 1 kilobase in length. Preferably, the probes are single copy probes that hybridize to a unique locus in haploid  
10 chromosome complement. Some exemplary restriction enzymes employed in RFLP mapping are *EcoRI*, *EcoRV*, and *SstI*. As used herein the term "restriction enzyme" includes reference to a composition that recognizes and, alone or in conjunction with another composition, cleaves at a specific nucleotide sequence.

The method of detecting an RFLP comprises the steps of (a) digesting genomic  
15 DNA of a plant with a restriction enzyme; (b) hybridizing a nucleic acid probe, under selective hybridization conditions, to a sequence of a polynucleotide of the present of said genomic DNA; (c) detecting therefrom a RFLP. Other methods of differentiating polymorphic (allelic) variants of polynucleotides of the present invention can be had by utilizing molecular marker techniques well known to those of skill in the art including  
20 such techniques as: 1) single stranded conformation analysis (SSCA); 2) denaturing gradient gel electrophoresis (DGGE); 3) RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6) allele-specific PCR. Other approaches based on the detection of mismatches between the two complementary DNA strands include  
25 clamped denaturing gel electrophoresis (CDGE); heteroduplex analysis (HA); and chemical mismatch cleavage (CMC). Thus, the present invention further provides a method of genotyping comprising the steps of contacting, under stringent hybridization conditions, a sample suspected of comprising a polynucleotide of the present invention with a nucleic acid probe. Generally, the sample is a plant sample, preferably, a sample  
30 suspected of comprising a maize polynucleotide of the present invention (e.g., gene, mRNA). The nucleic acid probe selectively hybridizes, under stringent conditions, to a



subsequence of a polynucleotide of the present invention comprising a polymorphic marker. Selective hybridization of the nucleic acid probe to the polymorphic marker nucleic acid sequence yields a hybridization complex. Detection of the hybridization complex indicates the presence of that polymorphic marker in the sample. In preferred  
5   embodiments, the nucleic acid probe comprises a polynucleotide of the present invention.

### UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA.  
10   Positive sequence motifs include translational initiation consensus sequences (Kozak, *Nucleic Acids Res* 15:8125 (1987)) and the 7-methylguanosine cap structure (Drummond *et al.*, *Nucleic Acids Res.* 13:7375 (1985)). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al.*, *Cell* 48:691 (1987)) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5'  
15   UTR (Kozak, *supra*, Rao *et al.*, *Mol. and Cell. Biol.* 8:284 (1988)). Accordingly, the present invention provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to  
20   alter translational efficiency and/or to optimize the coding sequence for expression in a desired host such as to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin  
25   Genetics Computer Group (see Devereaux *et al.*, *Nucleic Acids Res.* 12:387-395 (1984)) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present invention. The number of polynucleotides that can be used to determine a codon usage frequency can be any integer from 1 to the number of  
30   polynucleotides of the present invention as provided herein. Optionally, the

polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50, or 100.

### **Sequence Shuffling**

5           The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom. Sequence shuffling is described in PCT Publication No. WO 96/19256. See also, Zhang, J.-H., *et al. Proc. Natl. Acad. Sci. USA* 94:4504-4509 (1997). Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired  
10   characteristic, which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides which comprise sequence regions, which have substantial identity and can be homologously recombined in vitro or in vivo. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or  
15   advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a  
20   gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be a decreased  $K_m$  and/or increased  $K_{cat}$  over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled  
25   wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140%, or at least 150% of the wild-type value.

### **Generic and Consensus Sequences**

30           Polynucleotides and polypeptides of the present invention further include those having: (a) a generic sequence of at least two homologous polynucleotides or polypeptides, respectively, of the present invention; and, (b) a consensus sequence of at

least three homologous polynucleotides or polypeptides, respectively, of the present invention. The generic sequence of the present invention comprises each species of polypeptide or polynucleotide embraced by the generic polypeptide or polynucleotide, sequence, respectively. The individual species encompassed by a polynucleotide having  
5 an amino acid or nucleic acid consensus sequence can be used to generate antibodies or produce nucleic acid probes or primers to screen for homologs in other species, genera, families, orders, classes, phylums, or kingdoms. For example, a polynucleotide having a consensus sequence from a gene family of *Zea mays* can be used to generate antibody or nucleic acid probes or primers to other *Gramineae* species such as wheat, rice, or  
10 sorghum. Alternatively, a polynucleotide having a consensus sequence generated from orthologous genes can be used to identify or isolate orthologs of other taxa. Typically, a polynucleotide having a consensus sequence will be at least 9, 10, 15, 20, 25, 30, or 40 amino acids in length, or 20, 30, 40, 50, 100, or 150 nucleotides in length. As those of skill in the art are aware, a conservative amino acid substitution can be used for amino  
15 acids, which differ amongst aligned sequence but are from the same conservative amino substitution group as discussed above. Optionally, no more than 1 or 2 conservative amino acids are substituted for each 10 amino acid length of consensus sequence.

Similar sequences used for generation of a consensus or generic sequence include any number and combination of allelic variants of the same gene, orthologous, or  
20 paralogous sequences as provided herein. Optionally, similar sequences used in generating a consensus or generic sequence are identified using the BLAST algorithm's smallest sum probability (P(N)). Various suppliers of sequence-analysis software are listed in chapter 7 of *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds. Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John  
25 Wiley & Sons, Inc. (Supplement 30). A polynucleotide sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, or 0.001, and most preferably less than about 0.0001, or 0.00001. Similar  
30 polynucleotides can be aligned and a consensus or generic sequence generated using multiple sequence alignment software available from a number of commercial suppliers such as the Genetics Computer Group's (Madison, WI) PILEUP software, Vector NTI's

(North Bethesda, MD) ALIGNX, or Genecode's (Ann Arbor, MI) SEQUENCER.

Conveniently, default parameters of such software can be used to generate consensus or generic sequences.

Although the present invention has been described in some detail by way of  
5 illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practices within the scope of the appended claims.

#### EXAMPLE 1

10 This example describes the method of finding proteins that interact with Rar1-Interactors using a Yeast Two-Hybrid System.

The method used to identify proteins that interact with Rar1 was a HybriZap®  
Two-Hybrid Predigested Vector/Gigapack Cloning KitXR Library Construction Kit and a  
15 directional cDNA Synthesis Kit from Stratagene (LaJolla, California). ZmRar1 (SEQ ID NOS: 4 and 5) was subcloned into pBDGAL4 in-frame with the GAL4 binding domain. The construct was amplified in One-Shot cells, genotype TOP10, from Invitrogen (Carlsbad, California) and isolated. The construct was partially digested with EcoRI, blunted with pfu and religated (3243:20, 3243:22). The religated construct was  
20 transformed into One-Shot cells. Colonies containing pBDZmRar1 were recovered, cultured, minipreped and sequenced.

Yeast strain YRG-2 (Stratagene) was transformed with the pBDZmRar1 plasmid using the lithium acetate protocol of Gietz et al., *Nucl Acids Res* 20(6):1425 (1992). A yeast selective (SD) media was prepared according to the HybriZap® manual  
25 (Stratagene) with sorbitol omitted and dropout powders (Clonetech) substituted for the indicated dropout solution. SD broths and agars lacking Trp or Leu were used for selection of cells containing pBDZmRar1. SD agars lacking Trp and Leu were used to quantify total double-transformants.

A cDNA library was made from mRNA isolated from maize cells. The maize  
30 cells were treated with water or  $1 \times 10^6$  spores/ml of *Fusarium moniliforme*. Cells were harvested 2 and 6 hours after treatment. Total RNA was isolated using Tri-Reagent™ and mRNA was isolated using PolyAtract™ (Promega). Zap-cDNA synthesis kit

(Stratagene) was used to prepare cDNA, which was cloned into HybriZap® (Stratagene). The primary library was amplified and phagemid was excised from the secondary library. The phagemid prep was amplified in XL0LR cells and purified (Qiagen) to prepare library DNA for transformation into yeast. All library manipulations were performed  
5 according to the HybriZap® manual.

The library DNA was transformed into pBDZmRar1-containing yeast cells using the method of Geitz and Schiestl, 1995 (Geitz and Schiestl, *Methods in Molecular and Cellular Biology* 5:255-269 (1995)). Approximately  $2.6 \times 10^6$  transformants were screened on minimal medium lacking Trp, Leu, and His. SD agars and broths lacking  
10 Trp, Leu, and His were used to select for interaction or recover interactor plasmids. Triple-dropout SD agars were sometimes supplemented with 3 aminotriazol (3AT, Sigma) at 10 or 25 mM to increase the stringency of the His selection.

No colonies grew on 50 mM 3AT, five colonies grew on 25 mM 3 AT, eleven colonies grew on 10 mM 3AT, and sixteen colonies grew on media without 3AT. Only  
15 colony number 6 and number 16 turned blue in the presence of Xgal. Library plasmids were isolated from the two yeast colonies containing the putative interactors (Hoffman and Winston, *Gene* 57:267-272 (1987)). The plasmid prep was transformed into chemically-competent *E. coli* (DH5α, Life Technologies) that were plated on carbenicillin-containing plates to select for the activation-domain plasmid. Resulting  
20 colonies were cultured, minipreped, and submitted for sequencing. Sequences were assembled using Sequencer 3.0 (GeneCodes) from sequence runs using both vector primers and internal primers.

Both interactors were verified by co-transformation with pBDZmRar1 and with pBDGal4CAM, which is the bait vector without insert. Both interactors caused  
25 expression of both reporter genes (HIS3 and lacZ) when co-transformed with pBDZmRar1 and no reporter gene expression when co-transformed with the bait vector alone.

## EXAMPLE 2

### Isolation of the Full Length Sequence of the Maize Rar1-Interactors

Total RNA was isolated from corn tissues with TRIzol Reagent (Life Technology Inc. Gaithersburg, MD) using a modification of the guanidine isothiocyanate/acid-phenol  
5 procedure described by Chomczynski and Sacchi (Chomczynski, P., and Sacchi, N. *Anal. Biochem.* 162, 156 (1987)). In brief, plant tissue samples were pulverized in liquid nitrogen before the addition of the TRIzol Reagent, and then were further homogenized with a mortar and pestle. Addition of chloroform followed by centrifugation was conducted for separation of an aqueous phase and an organic phase. The total RNA was  
10 recovered by precipitation with isopropyl alcohol from the aqueous phase.

The selection of poly(A)<sup>+</sup> RNA from total RNA was performed using PolyATact system (Promega Corporation, Madison WI). In brief, biotinylated oligo(dT) primers were used to hybridize to the 3' poly(A) tails on mRNA. The hybrids were captured using streptavidin coupled to paramagnetic particles and a magnetic separation stand.  
15 The mRNA was washed at high stringent condition and eluted by RNase-free deionized water.

cDNA synthesis was performed and unidirectional cDNA libraries were constructed using the SuperScript Plasmid System (Life Technology Inc. Gaithersburg, MD). The first strand of cDNA was synthesized by priming an oligo(dT) primer  
20 containing a Not I site. The reaction was catalyzed by SuperScript reverse Transcriptase II at 45°C. The second strand of cDNA was labeled with alpha-<sup>32</sup>P-dCTP and a portion of the reaction was analyzed by agarose gel electrophoresis to determine cDNA sizes. cDNA molecules smaller than 500 base pairs and unligated adaptors were removed by Sephacryl-S400 chromatography. The selected cDNA molecules were ligated into a  
25 pSPORT1 vector between the *NotI* and *Sall* sites.

Individual colonies were picked and DNA was prepared either by PCR with M13 forward primers and M13 reverse primers, or by plasmid isolation. All the cDNA clones were sequenced using M13 reverse primers.

cDNA libraries subjected to the subtraction procedure were plated out on 22 x 22  
30 cm<sup>2</sup> agar plate at density of about 3,000 colonies per plate. The plates were incubated in a 37°C incubator for 12-24 hours. Colonies were picked into 384-well plates by a robot

colony picker, Q-bot (GENETIX Limited). These plates were incubated overnight at 37°C.

Once sufficient colonies were picked, they were pinned onto 22 x 22 cm<sup>2</sup> nylon membranes using Q-bot. Each membrane contained 9,216 colonies or 36,864 colonies.

- 5 These membranes were placed onto agar plate with appropriate antibiotic. The plates were incubated at 37°C for overnight.

After colonies were recovered on the second day, these filters were placed on filter paper prewetted with denaturing solution for four minutes, then were incubated on top of a boiling water bath for additional four minutes. The filters were then placed on  
10 filter paper prewetted with neutralizing solution for four minutes. After excess solution was removed by placing the filters on dry filter papers for one minute, the colony site of the filters were placed into Proteinase K solution, incubated at 37°C for 40-50 minutes. The filters were placed on dry filter papers to dry overnight. DNA was then cross-linked to nylon membrane by UV light treatment.

- 15 Colony hybridization was conducted as described by Sambrook, J., Fritsch, E.F. and Maniatis, T., (in Molecular Cloning: A laboratory Manual, 2<sup>nd</sup> Edition). The following probes were used in colony hybridization:

1. First strand cDNA from the same tissue as the library was made from to remove the most redundant clones.
- 20 2. 48-192 most redundant cDNA clones from the same library based on previous sequencing data.
3. 192 most redundant cDNA clones in the entire corn sequence database.
4. A Sal-A20 oligo nucleotide TCG ACC CAC GCG TCC GAA AAA AAA AAA AAA AAA AAA, (SEQ ID NO: 6) removes clones containing a poly A tail but no  
25 cDNA.
5. cDNA clones derived from rRNA.

The image of the autoradiography was scanned into computer and the signal intensity and cold colony addresses of each colony was analyzed, re-arraying of cold-colonies from 384 well plates to 96 well plates was conducted using Q-bot.

30

### EXAMPLE 3

This example describes identification of the gene from a computer homology search.

Gene identities can be determined by conducting BLAST (Basic Local Alignment  
5 Search Tool; Altschul, S. F., et al., (1993) J. Mol. Biol. 215:403-410; see also  
www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to  
sequences contained in the BLAST "nr" database (comprising all non-redundant  
GenBank CDS translations, sequences derived from the 3-dimensional structure  
10 Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein  
sequence database, EMBL, and DDBJ databases). The cDNA sequences are analyzed for  
similarity to all publicly available DNA sequences contained in the "nr" database using  
the BLASTN algorithm. The DNA sequences are translated in all reading frames and  
compared for similarity to all publicly available protein sequences contained in the "nr"  
15 database using the BLASTX algorithm (Gish, W. and States, D. J. *Nature Genetics*  
3:266-272 (1993)) provided by the NCBI. In some cases, the sequencing data from two  
or more clones containing overlapping segments of DNA are used to construct  
contiguous DNA sequences.

Sequence alignments and percent identity calculations can be performed using the  
Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR  
20 Inc., Madison, WI). Multiple alignment of the sequences can be performed using the  
Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the  
default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default  
parameters for pairwise alignments using the Clustal method are KTUPLE 1, GAP  
PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

25 Using the partial sequence identified by the yeast two-hybrid system, full length  
clones were identified from the maize cDNA library as described above. The first maize  
Rar1-interactor, RarInt1, polynucleotide sequence can be found in SEQ ID NO: 1.  
RarInt1 is a strong interactor with ZmRar1 and shows homology to a wheat heat-shock  
protein 80 (Accession No. Z11920), with BLASTn segment scores of 818 and 386 with a  
30 0.0 expectation and 87% identity by comparison of the polynucleotides with GAP



analysis. RarInt1 also shows homology to a rice heat shock protein 82 (Accession No. Z11920), by 68% identity by GAP analysis.

The second maize Rar1-interactor, RarInt2, is a moderate interactor and shows homology to the rice SGT1 gene (Accession No. AF192467). The partial polynucleotide  
5 sequence for RarInt2 can be found in SEQ ID NOS: 2 and 3. The BLASTn segment scores for comparing RarInt2 to the rice SGT1 gene are 121 (expectation  $2 \times 10^{-61}$ ), 40 (expectation of  $9 \times 10^{-13}$ ), and 37 (expectation of  $6 \times 10^{-11}$ ). Comparison of SEQ ID NOS: 2 and 3 with the rice SGT1 gene by GAP analysis, shows a 78% identity with SEQ ID NO: 2 and a 40% identity with SEQ ID NO: 3.

10 The full length nucleotide sequence of RarInt2 was obtained by sequencing the clone deposited with the Patent Depository of the American Type Culture Collection (ATCC), Manassas, Virginia, on February 21, 2001 and assigned Patent Deposit No. PTA-3031.

The full polynucleotide sequence for RarInt2 can be found in SEQ ID NO:7; the  
15 coding sequence for RarInt2 (nucleotides 82-1167 of SEQ ID NO:7) is set forth in SEQ ID NO:9. Comparison of SEQ ID NO:9 with the nucleotide sequence of the SGT1 genes of barley (Accession No. AF439974), rice (Accession No. AF192467), and *Arabidopsis thaliana* (SGT1a, Accession No. AF439975; SGT1b, Accession No. AF439976) by GAP analysis shows 82.825%, 82.222%, 64.299%, and 63.138% identity, respectively.

20 The polypeptide encoded by SEQ ID NOS:7 and 9 is set forth in SEQ ID NO:8. Comparison of SEQ ID NO:8 with the SGT1 genes of barley (Accession No. AAL33610), rice (Accession Nos. BAB19060 and AAF18438), and *Arabidopsis thaliana* SGT1a (Accession No. AAL33611) by GAP analysis, shows a 82%, 84%, 83%, and 61% identity, respectively.

25

#### EXAMPLE 4

##### Transformation and Regeneration of Transgenic Plants

Immature maize embryos from greenhouse donor plants are bombarded with a plasmid containing the Rar1-interactor sequences of the present invention operably linked  
30 to a ubiquitin promoter and the selectable marker gene PAT (Wohlleben *et al.* (1988) *Gene* 70:25-37), which confers resistance to the herbicide Bialaphos. Alternatively, the

selectable marker gene is provided on a separate plasmid. Transformation is performed as follows. Media recipes follow below.

#### Preparation of Target Tissue

5 The ears are husked and surface sterilized in 30% Clorox bleach plus 0.5% Micro detergent for 20 minutes, and rinsed two times with sterile water. The immature embryos are excised and placed embryo axis side down (scutellum side up), 25 embryos per plate, on 560Y medium for 4 hours and then aligned within the 2.5-cm target zone in preparation for bombardment.

#### Preparation of DNA

10 This plasmid DNA containing the Rar1-interactor plus plasmid DNA containing a PAT selectable marker is precipitated onto 1.1  $\mu\text{m}$  (average diameter) tungsten pellets using a  $\text{CaCl}_2$  precipitation procedure as follows:

100  $\mu\text{l}$  prepared tungsten particles in water  
15 10  $\mu\text{l}$  (1  $\mu\text{g}$ ) DNA in Tris EDTA buffer (1  $\mu\text{g}$  total DNA)  
100  $\mu\text{l}$  2.5 M  $\text{CaCl}_2$   
10  $\mu\text{l}$  0.1 M spermidine

Each reagent is added sequentially to the tungsten particle suspension, while  
20 maintained on the multitube vortexer. The final mixture is sonicated briefly and allowed to incubate under constant vortexing for 10 minutes. After the precipitation period, the tubes are centrifuged briefly, liquid removed, washed with 500 ml 100% ethanol, and centrifuged for 30 seconds. Again the liquid is removed, and 105  $\mu\text{l}$  100% ethanol is added to the final tungsten particle pellet. For particle gun bombardment, the  
25 tungsten/DNA particles are briefly sonicated and 10  $\mu\text{l}$  spotted onto the center of each macrocarrier and allowed to dry about 2 minutes before bombardment.

#### Particle Gun Treatment

The sample plates are bombarded at level #4 in particle gun #HE34-1 or #HE34-  
2. All samples receive a single shot at 650 PSI, with a total of ten aliquots taken from  
30 each tube of prepared particles/DNA.

### Subsequent Treatment

Following bombardment, the embryos are kept on 560Y medium for 2 days, then transferred to 560R selection medium containing 3 mg/liter Bialaphos, and subcultured every 2 weeks. After approximately 10 weeks of selection, selection-resistant callus clones are transferred to 288J medium to initiate plant regeneration. Following somatic embryo maturation (2-4 weeks), well-developed somatic embryos are transferred to medium for germination and transferred to the lighted culture room. Approximately 7-10 days later, developing plantlets are transferred to 272V hormone-free medium in tubes for 7-10 days until plantlets are well established. Plants are then transferred to inserts in flats (equivalent to 2.5" pot) containing potting soil and grown for 1 week in a growth chamber, subsequently grown an additional 1-2 weeks in the greenhouse, then transferred to classic 600 pots (1.6 gallon) and grown to maturity. Plants are monitored and scored for and altered level of expression of the Rar1-interactor sequence of the invention. Alternatively, the Rar1-interactor activity can be assayed (i.e., enhance disease resistance).

### Bombardment and Culture Media

Bombardment medium (560Y) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 120.0 g/l sucrose, 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-I H<sub>2</sub>O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature). Selection medium (560R) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 30.0 g/l sucrose, and 2.0 mg/l 2,4-D (brought to volume with D-I H<sub>2</sub>O following adjustment to pH 5.8 with KOH); 3.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 0.85 mg/l silver nitrate and 3.0 mg/l bialaphos (both added after sterilizing the medium and cooling to room temperature).

Plant regeneration medium (288J) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H<sub>2</sub>O) (Murashige and Skoog (1962) *Physiol. Plant.* 15:473), 100 mg/l myo-inositol, 0.5 mg/l

zeatin, 60 g/l sucrose, and 1.0 ml/l of 0.1 mM abscisic acid (brought to volume with polished D-I H<sub>2</sub>O after adjusting to pH 5.6); 3.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 1.0 mg/l indoleacetic acid and 3.0 mg/l bialaphos (added after sterilizing the medium and cooling to 60°C). Hormone-free medium (272V) comprises  
5 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g/l nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H<sub>2</sub>O), 0.1 g/l myo-inositol, and 40.0 g/l sucrose (brought to volume with polished D-I H<sub>2</sub>O after adjusting pH to 5.6); and 6 g/l bacto-agar (added after bringing to volume with polished D-I H<sub>2</sub>O), sterilized and cooled to 60° C.

10

## EXAMPLE 5

### *Agrobacterium*-mediated Transformation of Maize

For *Agrobacterium*-mediated transformation of maize with a Rar1-interactor polynucleotide operably linked to ubiquitin promoter, preferably the method of Zhao is  
15 employed (U.S. Patent No. 5,981,840, and PCT patent publication WO98/32326; the contents of which are hereby incorporated by reference). Briefly, immature embryos are isolated from maize and the embryos contacted with a suspension of *Agrobacterium*, where the bacteria are capable of transferring the Rar1-interactor nucleotide sequences to at least one cell of at least one of the immature embryos (step 1: the infection step). In  
20 this step the immature embryos are preferably immersed in an *Agrobacterium* suspension for the initiation of inoculation. The embryos are co-cultured for a time with the *Agrobacterium* (step 2: the co-cultivation step). Preferably the immature embryos are cultured on solid medium following the infection step. Following this co-cultivation period an optional "resting" step is contemplated. In this resting step, the embryos are  
25 incubated in the presence of at least one antibiotic known to inhibit the growth of *Agrobacterium* without the addition of a selective agent for plant transformants (step 3: resting step). Preferably the immature embryos are cultured on solid medium with antibiotic, but without a selecting agent, for elimination of *Agrobacterium* and for a resting phase for the infected cells. Next, inoculated embryos are cultured on medium  
30 containing a selective agent and growing transformed callus is recovered (step 4: the selection step). Preferably, the immature embryos are cultured on solid medium with a

selective agent resulting in the selective growth of transformed cells. The callus is then regenerated into plants (step 5: the regeneration step), and preferably calli grown on selective medium are cultured on solid medium to regenerate the plants.

5

## EXAMPLE 6

### Soybean Embryo Transformation

Soybean embryos are bombarded with a plasmid containing the Rar1-interactor polynucleotide operably linked to an Scp1 promoter (U.S. Patent No. 6,072,050) as follows. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from  
10 surface-sterilized, immature seeds of the soybean cultivar A2872, are cultured in the light or dark at 26°C on an appropriate agar medium for six to ten weeks. Somatic embryos producing secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos that multiplied as early, globular-staged embryos, the suspensions are maintained as described below.

15 Soybean embryogenic suspension cultures can be maintained in 35 ml liquid media on a rotary shaker, 150 rpm, at 26° C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 ml of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method  
20 of particle gun bombardment (Klein *et al.* (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A Du Pont Biolistic PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene that can be used to facilitate soybean transformation is a transgene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell *et al.*  
25 (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz *et al.* (1983) *Gene* 25:179-188), and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The expression cassette comprising the Rar1-interactor sequence operably linked to the Scp1 promoter can be isolated as a restriction fragment. This fragment can  
30 then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50  $\mu$ l of a 60 mg/ml 1  $\mu$ m gold particle suspension is added (in order): 5  $\mu$ l DNA (1  $\mu$ g/ $\mu$ l), 20  $\mu$ l spermidine (0.1 M), and 50  $\mu$ l  $\text{CaCl}_2$  (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400  $\mu$ l 70% ethanol and resuspended in 40  $\mu$ l of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five microliters of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi, and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post-bombardment with fresh media containing 50 mg/ml hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post-bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### EXAMPLE 7

##### Sunflower Meristem Tissue Transformation

Sunflower meristem tissues are transformed with an expression cassette containing the Rar1-interactor polynucleotide operably linked to a ubiquitin promoter as follows (see also European Patent Number EP 0 486233, herein incorporated by

reference, and Malone-Schoneberg *et al.* (1994) *Plant Science* 103:199-207). Mature sunflower seed (*Helianthus annuus* L.) are dehulled using a single wheat-head thresher. Seeds are surface sterilized for 30 minutes in a 20% Clorox bleach solution with the addition of two drops of Tween 20 per 50 ml of solution. The seeds are rinsed twice with  
5 sterile distilled water.

Split embryonic axis explants are prepared by a modification of procedures described by Schrammeijer *et al.* (Schrammeijer *et al.* (1990) *Plant Cell Rep.* 9:55-60). Seeds are imbibed in distilled water for 60 minutes following the surface sterilization procedure. The cotyledons of each seed are then broken off, producing a clean fracture at  
10 the plane of the embryonic axis. Following excision of the root tip, the explants are bisected longitudinally between the primordial leaves. The two halves are placed, cut surface up, on GBA medium consisting of Murashige and Skoog mineral elements (Murashige *et al.* (1962) *Physiol. Plant.*, 15: 473-497), Shepard's vitamin additions (Shepard (1980) in *Emergent Techniques for the Genetic Improvement of Crops*  
15 (University of Minnesota Press, St. Paul, Minnesota), 40 mg/l adenine sulfate, 30 g/l sucrose, 0.5 mg/l 6-benzyl-aminopurine (BAP), 0.25 mg/l indole-3-acetic acid (IAA), 0.1 mg/l gibberellic acid (GA<sub>3</sub>), pH 5.6, and 8 g/l Phytagar.

The explants are subjected to microprojectile bombardment prior to *Agrobacterium* treatment (Bidney *et al.* (1992) *Plant Mol. Biol.* 18:301-313). Thirty to  
20 forty explants are placed in a circle at the center of a 60 X 20 mm plate for this treatment. Approximately 4.7 mg of 1.8 mm tungsten microprojectiles are resuspended in 25 ml of sterile TE buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0) and 1.5 ml aliquots are used per bombardment. Each plate is bombarded twice through a 150 mm nytex screen placed 2 cm above the samples in a PDS 1000® particle acceleration device.

25 Disarmed *Agrobacterium tumefaciens* strain EHA105 is used in all transformation experiments. A binary plasmid vector comprising the expression cassette that contains the Rar1-interactor gene operably linked to the constitutive promoter is introduced into *Agrobacterium* strain EHA105 via freeze-thawing as described by Holsters *et al.* (1978) *Mol. Gen. Genet.* 163:181-187. This plasmid further comprises a kanamycin selectable  
30 marker gene (i.e., *nptII*). Bacteria for plant transformation experiments are grown overnight (28°C and 100 RPM continuous agitation) in liquid YEP medium (10 gm/l

yeast extract, 10 gm/l Bactopeptone, and 5 gm/l NaCl, pH 7.0) with the appropriate antibiotics required for bacterial strain and binary plasmid maintenance. The suspension is used when it reaches an OD<sub>600</sub> of about 0.4 to 0.8. The *Agrobacterium* cells are pelleted and resuspended at a final OD<sub>600</sub> of 0.5 in an inoculation medium comprised of  
5 12.5 mM MES pH 5.7, 1 gm/l NH<sub>4</sub>Cl, and 0.3 gm/l MgSO<sub>4</sub>.

Freshly bombarded explants are placed in an *Agrobacterium* suspension, mixed, and left undisturbed for 30 minutes. The explants are then transferred to GBA medium and co-cultivated, cut surface down, at 26°C and 18-hour days. After three days of co-cultivation, the explants are transferred to 374B (GBA medium lacking growth regulators  
10 and a reduced sucrose level of 1%) supplemented with 250 mg/l cefotaxime and 50 mg/l kanamycin sulfate. The explants are cultured for two to five weeks on selection and then transferred to fresh 374B medium lacking kanamycin for one to two weeks of continued development. Explants with differentiating, antibiotic-resistant areas of growth that have not produced shoots suitable for excision are transferred to GBA medium containing 250  
15 mg/l cefotaxime for a second 3-day phytohormone treatment. Leaf samples from green, kanamycin-resistant shoots are assayed for the presence of NPTII by ELISA and for the presence of transgene expression by monitoring and scoring for an altered level of expression of the Rar-1-interactor sequence of the invention. Alternatively, the Rar1-interactor activity can be assayed (i.e., enhanced disease resistance).

20 NPTII-positive shoots are grafted to Pioneer<sup>®</sup> hybrid 6440 *in vitro*-grown sunflower seedling rootstock. Surface sterilized seeds are germinated in 48-0 medium (half-strength Murashige and Skoog salts, 0.5% sucrose, 0.3% gelrite, pH 5.6) and grown under conditions described for explant culture. The upper portion of the seedling is removed, a 1 cm vertical slice is made in the hypocotyl, and the transformed shoot  
25 inserted into the cut. The entire area is wrapped with parafilm to secure the shoot. Grafted plants can be transferred to soil following one week of *in vitro* culture. Grafts in soil are maintained under high humidity conditions followed by a slow acclimatization to the greenhouse environment. Transformed sectors of T<sub>0</sub> plants (parental generation) maturing in the greenhouse are identified by NPTII ELISA and/or by Rar1-interactor  
30 activity analysis of leaf extracts while transgenic seeds harvested from NPTII-positive T<sub>0</sub>



plants are identified by Rar1-interactor activity analysis of small portions of dry seed cotyledon.

5       The above examples are provided to illustrate the invention but not to limit its  
scope. Other variants of the invention will be readily apparent to one of ordinary skill in  
the art and are encompassed by the appended claims. All publications, patents, and  
patent applications cited herein are indicative of the level of those skilled in the art to  
which this invention pertains. All publications, patents, and patent applications are  
hereby incorporated by reference to the same extent as if each individual publication or  
10   patent application was specifically and individually indicated to be incorporated by  
reference.